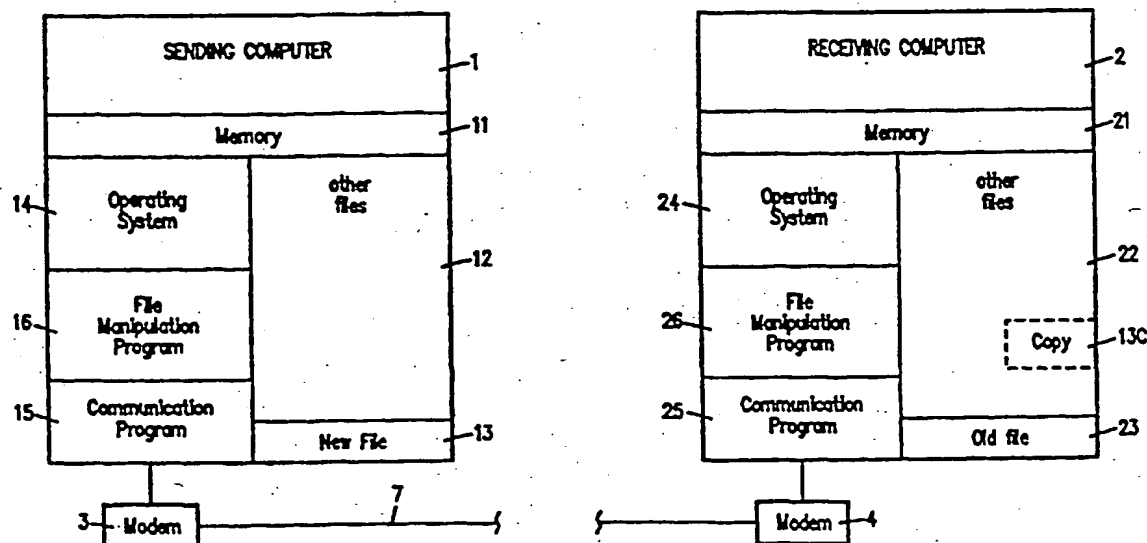




INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| | | |
|--|---|--|
| (51) International Patent Classification ⁶ : G06F 13/00 | A1 | (11) International Publication Number: WO 95/19003 (43) International Publication Date: 13 July 1995 (13.07.95) |
| (21) International Application Number: PCT/US94/14969 (22) International Filing Date: 27 December 1994 (27.12.94) (30) Priority Data: 08/176,955 3 January 1994 (03.01.94) US (71) Applicant: NORTON-LAMBERT CORP. [US/US]; 5209 Overpass Road, Building A, Santa Barbara, CA 93140 (US). , Inventors: HARLAN, Jim; 24000 Telegraph Road, Southfield, MI 48034 (US). THOMAS, Henry, Edward; 1211 Pearl Street, Ypsilanti, MI 48197 (US). (74) Agent: GALBI, Elmer; 13314 Vermeer Drive, Lake Oswego, OR 97035 (US). | (81) Designated States: CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i> | |

(54) Title: FILE TRANSFER METHOD AND APPARATUS USING HASH NUMBERS



(57) Abstract

The present invention facilitates and speeds the transmission of a copy of a new file (13) to a receiving computer (2) where the receiving computer (2) has an old file (23). The sending computer (1) does not know the status or content of the old file (23). As a preliminary step, the receiving computer divides the old file into segments, and calculates a hash number for each segment. The receiving computer (2) then transmits these hash numbers to the sending computer (1). The sending computer (1) examines each segment in the new file (13) to determine which, if any, segments in the new file (13) have hash numbers that match the hash numbers received from the receiving computer (2). The sending computer (1) sends to the receiving computer (2) those bytes from the new file (13) that are not part of any matching segment and an indication where matching segments fit into the new file (13). Finally, the receiving computer (2) constructs a copy of the new file (13C) from the bytes received and from the matching segments in the old file (23).

1 **FILE TRANSFER METHOD AND APPARATUS USING HASH NUMBERS**

2 **Field of the Invention:**

3 The present invention relates to electronic computers and
4 more particularly to the transfer of files between
5 computers.

6 **Background of the Invention:**

7 There are a wide variety of commercially available
8 computer programs which facilitate transferring files
9 between computers utilizing modems and telephone lines.
10 Among such commercially available programs are:
11 "Crosstalk" marketed by DCA Corp of Atlanta, Georgia;
12 "QModem" marketed by Forgin Inc. of Cedar Falls Iowa;
13 and, "Close-Up", marketed by Norton Lambert Corp of Santa
14 Barbara, CA.

15 The physical characteristics of normal telephone lines
16 limit the transmission speed which can be used to
17 transmit data over such lines. In order to shorten the
18 time required to transmit data, various data compression
19 and error correcting protocols are in widespread use.

20 It is known that when a file on a particular computer is,
21 being updated, the transmission time can be shorted by
22 merely transmitting information which relates to the
23 "differences" or the "delta" between the present file and

1 the previously transmitted file. The technique of only
2 transmitting the delta between two files is only
3 applicable in situations where the sending system knows
4 the state of the file at the receiving station.

5 The present invention provides a technique for rapidly
6 transmitting files between computers where the computer
7 receiving the information has a file stored thereon, but
8 where the sending computer does not know the state (i.e.
9 the exact contents) of the file at the receiving
10 computer.

11 Summary of the Invention:

12 The present invention facilitates the transmission of a
13 file (hereinafter referred to as the new file) from a
14 first computer to a second computer, in a situation where
15 the second computer has a file (hereinafter referred to
16 as the old file) but where the first computer does not
17 know the status or content of the old file. With the
18 present invention, as a preliminary step, the second
19 computer divides the old file into segments, and
20 calculates a hash number for each segment. The second
21 computer transmits these hash numbers to the first
22 computer. The first computer examines each possible
23 segment in the new file to determine which if any
24 segments in the new file have hash numbers which
25 correspond to the hash numbers received from the second
26 computer (such segments are hereinafter called matching

1 segments). The first computer sends to the second
2 computer those bytes from the new file that are not part
3 of any matching segment and an indication of which
4 matching segments fit into the new file. The second
5 computer constructs copy of the new file from the bytes
6 received and from the matching segments in the old file.

7 Brief Description of the Drawings:

8 Figure 1 is an overall block diagram of the computer
9 systems.

10 Figure 2 is a block diagram of the actions that take
11 place at the receiving computer.

12 Figure 3A is an example of a Segment Profile Table.

13 Figure 3B is a table giving an example of information
14 transmitted by the sending computer.

15 Figure 4 is a block diagram of the actions that take
16 place at the sending computer.

17 Detailed Description of a Preferred embodiment:

18 Two interconnected computers that can be used to practice
19 the invention are shown in Figure 1. A sending computer
20 1 is connected to a receiving computer 2 via modems 3 and
21 4 and a telephone line 7. A new file 13, is stored in
22 computer 1. The preferred embodiment of the invention

1 described herein can be used to transfer a copy of file
2 13 from computer 1 to computer 2. The copy of the file
3 13 which resides in computer 2 after the transfer
4 operation is designated 13C. It is shown in dotted lines
5 in Figure 1 since it is only present in computer 2 after
6 the transfer operation is complete.

7 Computer 1 has a conventional internal RAM memory 11
8 which has stored therein a number of programs and files.
9 It is noted that while various programs and files are
10 shown as being in the RAM memory 11 of computer 1, a
11 substantial part of these programs and files could
12 alternatively be stored on other types conventional
13 storage devices such as on magnetic disks. How the
14 various programs and files are stored is not particularly
15 relevant to the present invention and it can be in any
16 conventional manner.

17 As shown in Figure 1, the computer memory 11 includes an
18 operating system 14, a communication program 15, a file
19 manipulation program 16, a new file 13 and other files
20 12. The operating system 14 can for example be the DOS
21 operating system that is marketed by Microsoft
22 Corporation and the communications program 15 can be a
23 conventional communication program for a DOS type of
24 computer. The new file 13 is the file which computer 1
25 will transferred to computer 2 utilizing the present
26 invention. It is noted that as used herein the term

1 "transferring a file" should be understood as synonymous
2 with the more precise terms "transferring a copy of a
3 file". Furthermore as will be explained hereinafter in
4 alternative embodiments of the invention, the "file"
5 being transferred may merely be a designated string of
6 bytes and not a complete file in the sense that a DOS
7 file is a complete file..

8 The file manipulation program 16, and related program 26
9 in computer 2, are the programs which implement the main
10 parts of the present invention as hereinafter described.
11 The operations performed by these programs are shown in
12 block diagram form in Figures 2 and 4.

13 The computer 2 is substantially identical to the computer
14 1 and the components in computer 2, other than the files,
15 are identical. Computer 2 includes memory 21, operating
16 system 24, file manipulation program 26, communication
17 program 25 and a file designated "old file" 23.

18 It is noted that new file 13, and old file 23, are merely
19 illustrative of files that are typically stored on
20 personal computers and work stations. Typically a
21 computer will have many stored files and often the user
22 of a computer has a desire and need to transfer a file to
23 another computer. There are many existing programs and
24 protocols designed for this purpose. Many of these
25 protocols involve various types of compression..

1 Conventional communication programs 15 and 25 may or may
2 not use the conventional type of compression techniques
3 for transmitting data. The present invention relates to
4 the particular information which is transmitted in order
5 to transmit a complete file. The actual transmission
6 mechanism for transmitting the information may be
7 conventional.

8 The present invention takes a new and different approach
9 to the file transfer task. The present invention
10 recognizes and takes advantage of the fact that many
11 times when a file is being transferred from a first
12 computer to a second computer, there are files stored at
13 the second computer that are related in some way to the
14 file that is being transmitted. For example, new file 13
15 may be a updated version of the old file 23.
16 Alternatively, the new file 13 may be a file in the
17 format of a particular word processor document such a
18 WordPerfect. The document may for example by a contract.
19 The old file 23 may a WordPerfect document where the
20 document in file 23 may be an unrelated contract. The
21 present invention takes advantage of the fact that such
22 seemingly unrelated files may contain a substantial
23 number of identical stings of bytes. For example, some
24 of the similar bytes may be formatting information,
25 others similar bytes may be type fonts, others similar
26 bytes may be similar phrases that appear in the two
27 documents, etc. With the present invention similarities

1 between a new file (i.e. the file being transmitted) and
2 a document at the receiving computer are detected and
3 used to speed the transmission of
4 the new file.

5 The present invention also takes advantage that present
6 day modems 3 and 4 can operate in a full duplex mode
7 where information is transferred simultaneously in two
8 directions. The present invention utilizes bi-
9 directional transfer of information between computers to
10 speed the transfer of information in one direction.

11 The preferred embodiment described herein performs the
12 following major steps in order to transfer a copy of new
13 file 13 from computer 1 to computer 2.

14 a) A file stored on computer 2 is selected for
15 designation as old file 23.

16 b) Computer 2 divides the old file 23 into 128 byte
17 segments and calculates a hash number (e.g. a CRC
18 number) for each 128 byte segment of the old
19 file 23.

20 c) Computer 2 sends to computer 1 the hash numbers for
21 the segments of file 23. Computer 1 stores these hash
22 numbers in a Segment Profile Table (SPT).

23 d) Computer 1 calculates hash numbers for each possible
24 segment in the new file 13 and compares these hash
25 numbers to the hash numbers it has received from
26 computer 2. Segments in the new file which have

1 hash numbers that correspond to the hash number of a
2 segment in the old file are termed matching
3 segments.
4 e) Bytes in the new file that are not part of any
5 matching segment are transmitted from computer 1 to
6 computer 2.
7 f) Matching segments are not transmitted from computer 1
8 to computer 2. Instead the computer 1 sends computer 2
9 an indication that a particular matching segment
10 fits at a particular place in the file being
11 constructed at computer 2. The location where the
12 segments from the old file fit into the copy 13C
13 of the new file is evident from the sequence in
14 which bytes from the new file and segment
15 identifications are transmitted.
16 g) Computer 2 constructs a copy 13C of the new file 13
17 from the transmitted bytes and from matching
18 segments copied from the old file.

19 At the beginning of the transmission process after a file
20 on computer 2 has been designated as old file 23,
21 computer 1 has an empty SPT. At this point computer 1
22 begins calculating the hash number of the first segment
23 in new file 13. Since the SPT is empty the calculated
24 hash numbers will not match any hash number in the SPT.
25 Since no match is found, computer 1 sends the first byte
26 in the first segment of file 13 to computer 2. A new
27 byte from file 13 is then be added to the segment and a

1 new hash number calculated. The process repeats. At the
2 same time that computer 1 is examining segments of new
3 file 13, computer 23 is calculating hash numbers for
4 segments in old file 23 and transmitting these values to
5 computer 1 for storage in the SPT. As time progresses,
6 the SPT will contain more and more entries and matches
7 between hash numbers computer by computer 1 and
8 information in the SPT will begin to occur. When
9 computer 1 finds a matching segment (i.e. a segment that
10 has a hash number that matches one of the hash numbers in
11 the SPT), computer 1 merely sends computer 2 an
12 indication that a match has been detected and that
13 computer 2 should copy a particular segment from the old
14 file into the new file. When computer 1 finds a matching
15 segment and transmits this information to computer 2,
16 computer 2 will insert the matching segment from the old
17 file 23 immediately after the last byte that was received
18 from computer 1. Thus the sequence information is
19 received by computer 2 indicates where segments from the
20 old file 23 should be inserted in the copy 13C of the new
21 file.

22 Segments from old file 23 are identified in the SPT by
23 the offset of the first byte in the segment. Thus when
24 computer 1 sends to computer 2 the identification of a
25 segment (i.e. the offset of the first byte of a segment)
26 the computer 2 can identify which part of the old file 23
27 should be copied into the copy 13C of the new file 13

1 that is being constructed.

2 The type of hash number used in the preferred embodiment
3 of the invention described herein is the well-known
4 cyclical redundancy check (CRC) number. The manner of
5 calculating such numbers is well known. It is noted that
6 while as described herein, the calculated CRC number is
7 described as uniquely identifies a specific segment, as
8 is well known, this is only correct in a practical sense
9 and not in a strict mathematical sense. Errors can
10 occur in that the same CRC can sometimes be calculated
11 for two different segments. The number of such "errors"
12 is so low as to be negligible and as described herein the
13 CRC numbers are assumed to represent unique file
14 segments. The number of possible duplications (i.e.
15 errors) can be further reduced by using a longer CRC
16 polynomial. As is well know, for computational
17 efficiency, the length of the CRC is best chosen to match
18 the size of the computer's registers. In situations
19 where the length of the CRC is dictated by other
20 considerations, two concatenated CRC numbers can be used
21 to reduce the number of "errors". The manner of
22 calculating CRC numbers is well know and forms no part of
23 the present invention. Instead of using CRC numbers as
24 hash numbers, the other well known types of hash numbers
25 could be used.

26 Each of the above major steps will now be explained in

1 detail as will their purpose and how they are carried
2 out. Figures 2 and 4 are program flow diagrams showing
3 the operations that take place on computers 1 and 2.
4 Figure 3A is a diagram showing the information stored in
5 the SPT. Figure 3B is a table illustrating a
6 representative sequence of information that is
7 transmitted from computer 1 to computer 2.

8 It is noted that data can simultaneously flow in both
9 directions over communication line 7 from modem 3 to
10 modem 4 and from modem 4 to modem 3. That is line 7
11 operates in full duplex mode. Modems and communication
12 programs that handle duplex communication are well known.
13 The unique method and apparatus of the present invention
14 takes advantage of the ability to transfer data in the
15 reverse direction without slowing the transfer of data in
16 the forward direction in order to speed the transfer of
17 the file in the forward direction.

18 Two processes take place on receiving computer 2. First
19 a CRC number is calculated for each segment in the old
20 file. This first processes includes receiving the file
21 name and file type of the new file from computer 1 and
22 determining which file stored at computer 2 will be
23 designated as the old file. Second, computer 2 receives
24 bytes and segment identifications from computer 1 and a
25 copy 13C of the new file 13 is constructed. As shown in
26 Figure 2, the first process which takes place on computer

1 2 is indicated by blocks 201 to 206 and the second
2 process is indicated by blocks 210 to 214.

3 At the initiation of a file transfer operation, computer
4 1 send to computer 2, the file name and file type of the
5 file which will be transmitted (block 201). Computer 2
6 selects a file (block 202) which will be used as old file
7 23 based upon the following priorities:

8 1) Same file name and file type. If no such file, then,

9 2) Same file type and same file name except for two
10 characters. If no such file, then,

11 3) Same file type. If no such file, then,

12 4) Same file name. If no such file, then,

13 5) Longest available file.

14 It is noted that various alternative ways could also be
15 used to identify which file on computer 2 will be
16 designated as "old file" 23.

17 Next (block 203) the first 128 byte segment is read from
18 the file designated as old file 23 and the CRC number for
19 this segment is calculated (block 204). The calculated
20 CRC number is sent to computer 1 (block 205). It is
21 noted that as computer 1 receives a sequence of CRC
22 numbers, the offset of the beginning of the segment used
23 to calculate each CRC number is merely the offset of the
24 segment used to calculate the previous number increased
25 by 128. The operations in block 203, 204 and 205 repeat
26 until the end of file is detected at which time the old

1 file is closed (block 206). The operations indicated by
2 blocks 203, 204 and 205, in effect divide the old file 23
3 into segments, each 128 bytes long, a CRC is calculated
4 for each of these segments, and the CRC values are
5 transmitted to computer 1.

6 At the same time the operations indicated by blocks 202
7 to 206 are taking place, computer 2 is receiving
8 information from computer 1. This is indicated by block
9 210 to 214. Typical modern day computers can easily
10 handle such multitasking on a time shared basis.
11 Initially computer 1 sends computer 2 a series of bytes
12 from the new file 13 (block 210). As the process
13 progresses, identification of segments from old file 23
14 (block 212) will be received interspersed with bytes from
15 new file 13. How this occurs will be explained later
16 with reference to figures 3B and 4.

17 The receiving computer 2 builds the copy 13C of the new
18 file 13 (block 213) from the bytes received from computer
19 1 and from segments from old file 23 (when it receives a
20 segment identification). Finally an end of file
21 indication is received (block 214) and the process is
22 complete.

23 The CRC numbers that are sent from computer 2 to computer
24 1 are assembled in computer 1 in a Segment Profile Table
25 (SPT) such as that shown in Figure 3A. It is noted that

1 the numbers 0 to 8 shown in the first column of Figure 3A
2 are shown in Figure 3A merely for convenience of
3 illustration. In the actual SPT, since each segment is
4 128 bytes long, the segment identifications and the
5 actual SPT would be the offset of the beginning of each
6 segment (i. e. the numbers 0 to 8 multiplied by 128).

7 The operations which take place at computer 1 are shown
8 in Figure 4. As with the computer 2, two process proceed
9 simultaneously (i.e. in a multitasking mode) at computer
10 1. First there is a calculating and sending operation
11 indicated by blocks 401 to 410 and second there is the
12 receiving operation indicated by blocks 421 and 422.

13 When a file transfer is initiated the first step (block
14 401) involves transferring the file name and file type of
15 the new file 13 (i.e. the name of the file being
16 transferred) from computer 1 to computer 2. Next a file
17 pointer is set to "0" (block 402) and the first one
18 hundred and twenty eight bytes are read from new file 13
19 and the CRC of this first segment is calculated (block
20 403). The CRC so calculated is compared to the CRC
21 numbers in the SPT (block 405). When the operation
22 begins, no CRC numbers will as of yet been received from
23 computer 2 and the SPT will be empty, thus no match will
24 be found. The first byte in the segment will therefore
25 be sent to computer 2 (block 407), the effect of that one
26 byte on the CRC is subtracted from the calculated CRC

1 (block 408), the next byte read from the file, and a new
2 CRC calculated (block 409). If the end of the file has
3 not been reached (block 404), the new CRC will be
4 compared to the contents of the SPT and the operation
5 proceeds. If there is a match between the calculated CRC
6 number and a CRC number in the SPT, the segment number of
7 the file from the SPT will be sent to computer 2 (block
8 406) and an entirely new 128 byte segment will be read
9 from the file (block 403).

10

11 The above process continues until and end of file
12 indication is detected (block 404). When the end of file
13 indication is detected, the bytes remaining (which could
14 be up to 127 bytes) are sent to computer 2 (block 410).

15 As indicated by block 421, the computer 1 receives the
16 CRC numbers calculated as indicated in Figure 2.

17 Computer 1 uses these numbers (block 422) to build a
18 Segment Profile Table (SPT) as indicated in Figure 3A.

19 An example of the sequence in which information is
20 transmitted from computer 1 to computer 2 is given in
21 Figure 3A. The reference numbers in the first column of
22 figure 3A are merely for reference during the explanation
23 of the table. The second column gives the information
24 transmitted and the third column is merely an brief
25 explanation to facilitate and understanding of Figure 3A.

1 As indicated by line L1, the first information
2 transmitted from computer 1 to computer 2 is the file
3 name and file type of the new file 13. This information
4 is used by computer 2, to select a file which will be
5 used as old file 23. Next in the example shown, bytes 1-
6 57 of the file are transmitted. This indicates that for
7 the particular file in question the first fifty seven
8 times through the loop formed by blocks 403 to 405 in
9 Figure 4, no matching CRC was found in the SPT. Next as
10 indicated by line L3, block 405 in figure 4 determines
11 that the segment of the new file being examined has the
12 same CRC number as does segment 3 in the old file.
13 Computer 1 merely sends to computer 2 an indication that
14 after putting the first 57 bytes (i.e. line L2 from
15 Figure 3B) into the file 13C, computer 2 should copy
16 segment 3 from the old file 23 into the new file 13C.
17 The process then proceeds through lines L4 to L7 etc. It
18 is noted that line L7 in figure 3A shows that the same
19 segment from the old file can be used more than once in
20 constructing the new file 13C.

21 It is noted that the technique used to determine which
22 file on computer 2 is designated as the old file is not
23 particularly relevant to the present invention. Various
24 techniques could be used as an alternative to that shown
25 above. Naturally if the old file selected closely
26 resembles the new file 13, less bytes and more segments
27 identifications could be transmitted by computer 1

1 thereby reducing the transmission time.

2 It is also noted that the segments used in the above
3 described preferred embodiment are 128 bytes long.
4 Longer or shorter segments could be selected depending
5 upon the particular nature of the files being
6 transmitted. Furthermore the segment length could be
7 made dependent on various factors such as whether there
8 is a file in computer 2 with the same file name and file
9 type as the file in the sending computer or the file type
10 of the file begin transferred.

11 In the above described preferred embodiment, the
12 information in the SPT is only used to transmit one file,
13 herein designated new file 13. It is noted that the SPT
14 from each transmission operation can be saved such that
15 subsequent transmissions have at their disposal a large
16 SPT which defines segments in a plurality of files on the
17 receiving computer. Similar segments which could be
18 identified in a number of different files on the
19 receiving computer could then be used to speed the
20 transmission of one new file. That is the copy of the
21 new file would be made from segments which are identified
22 in a number of files on the receiving computer.

23 While in the preferred embodiment described above, a file
24 of the type used in DOS based computers was transferred
25 from computer 1 to computer 2, it should be understood

1 that in alternative embodiments, the invention could be
2 used to transfer other types of "files" between
3 computers. For example the invention could be used to
4 transfer a particular string of bytes from one computer
5 to a second computer. Thus, it should be understood that
6 the present invention can be used to transfer any string
7 of bytes (herein termed a "file") from one computer to a
8 second computer.

9 It is noted that herein computer 2 divides file 23 into
10 segments and calculates a CRC for each such segment,
11 while computer 1, calculates a CRC for each possible
12 segment, i.e each 128 byte segment following each byte in
13 the file. It is noted that in alternative embodiments,
14 computer two could also calculate hash numbers for
15 segments starting at various points in the file or
16 computer 1 could calculate CRC numbers by first dividing
17 the file into fixed length segments and later going back
18 and re-dividing the file into segments starting at
19 different places in the file.

20

21 While the invention has been described with reference to
22 a preferred embodiment thereof, it will be understood
23 that the alternatives mentioned and various other
24 alternatives could be chosen without departing from the
25 spirit and scope of the invention. The invention is
26 limited solely by the limitations in the appended claims.

1 We claim:

2 1) A method of transferring a particular file from a
3 first computer to a second computer, said second computer
4 having therein a second file, said method comprising the
5 steps of:
6 a) analyzing said second file and generating a hash
7 number for segments thereof,
8 b) transferring said hash numbers to said first computer
9 and storing them in a table,
10 d) analyzing said particular file to determine segments
11 thereof that have hash numbers corresponding to hash
12 numbers in said table, segments in said particular
13 file which have hash numbers corresponding to hash
14 numbers in said table comprising a first set of
15 segments,
16 e) sending to the second computer those parts of the
17 first file which are not part of any segment in said
18 first set of segments, and sending to the
19 second computer an indication of the segments
20 that are in said first set of segments, and
21 f) combining at said second computer the parts of said
22 particular file that were transmitted with
23 designated parts of the second file to construct a
24 replica of said particular file.

25 2) The method recited in claim 1 wherein said table is a
26 Segment Profile Table having a list of segments and an
27 indication of the hash number for each segment in the

1 list of segments.

2 3) The method recited in claim 1 wherein said hash
3 numbers are cyclical redundancy check numbers.

4 4) The method recited in claim 1 wherein said particular
5 file is a new file.

6 5) The method recited in claim 1 including the step of
7 sending the name of said particular file from said
8 first computer to said second computer.

9 6) The method recited in claim 5 wherein said second
10 computer has stored therein a plurality of files and
11 including the step of selecting a file at said second
12 computer which forms said second file.

13 7) The method recited in claim 6 wherein said selection
14 is based on the name of said particular file which was
15 sent from said first computer to said second computer.

- 1 8) A system for transferring a particular file from a
2 first computer to a second computer, said second computer
3 having therein a second file, said system comprising the
4 steps of:
- 5 a) means for analyzing said second file and generating a
6 hash number for each segment thereof,
- 7 b) means for transferring said hash numbers to said first
8 computer and storing them in a table,
- 9 d) means for analyzing said particular file to determine
10 segments thereof that have hash numbers
11 corresponding to hash numbers in said table,
12 segments in said particular file which have hash
13 numbers corresponding to hash numbers in said table
14 comprising a first set of segments,
- 15 e) means for sending to the second computer those parts
16 of the first file which are not part of any segment in
17 said first set of segments, and sending to the
18 second computer an indication of the segments that
19 are in said first set of segments, and
- 20 f) means for combining at said second computer the parts
21 of said particular file that were transmitted with
22 designated parts of said second file to construct a
23 replica of the said particular file.
- 24 9) The system recited in 8 wherein said hash numbers are
25 cyclical redundancy check (CRC) numbers.

1 10) A system for transferring a first string of bytes
2 from a first computer to a second computer, said
3 second computer having a second string of bytes
4 stored thereon, the string stored on said
5 second computer being devisable into segments,
6 means for transferring to said first computer a plurality
7 of hash numbers calculated from segments in said second
8 string of bytes, each hash number uniquely
9 identifying the content of the segment from which it
10 was calculated,
11 means for identifying a set of segments in said first
12 string of bytes which have hash numbers corresponding
13 to hash numbers transferred from said second
14 computer to said first computer,
15 means for transferring from said first computer to said
16 second computer the portions of said particular file
17 which are not in any segments in said set of
18 segments, and an indication of which segments from
19 said second set of bytes has a corresponding segment
20 in said first string of bytes,
21 means in said second computer for forming a third string
22 of bytes from the segments in the string of bytes on
23 said second computer identified in said set of
24 segments, and from the parts of said first string of
25 bytes which are transmitted to said second computer.

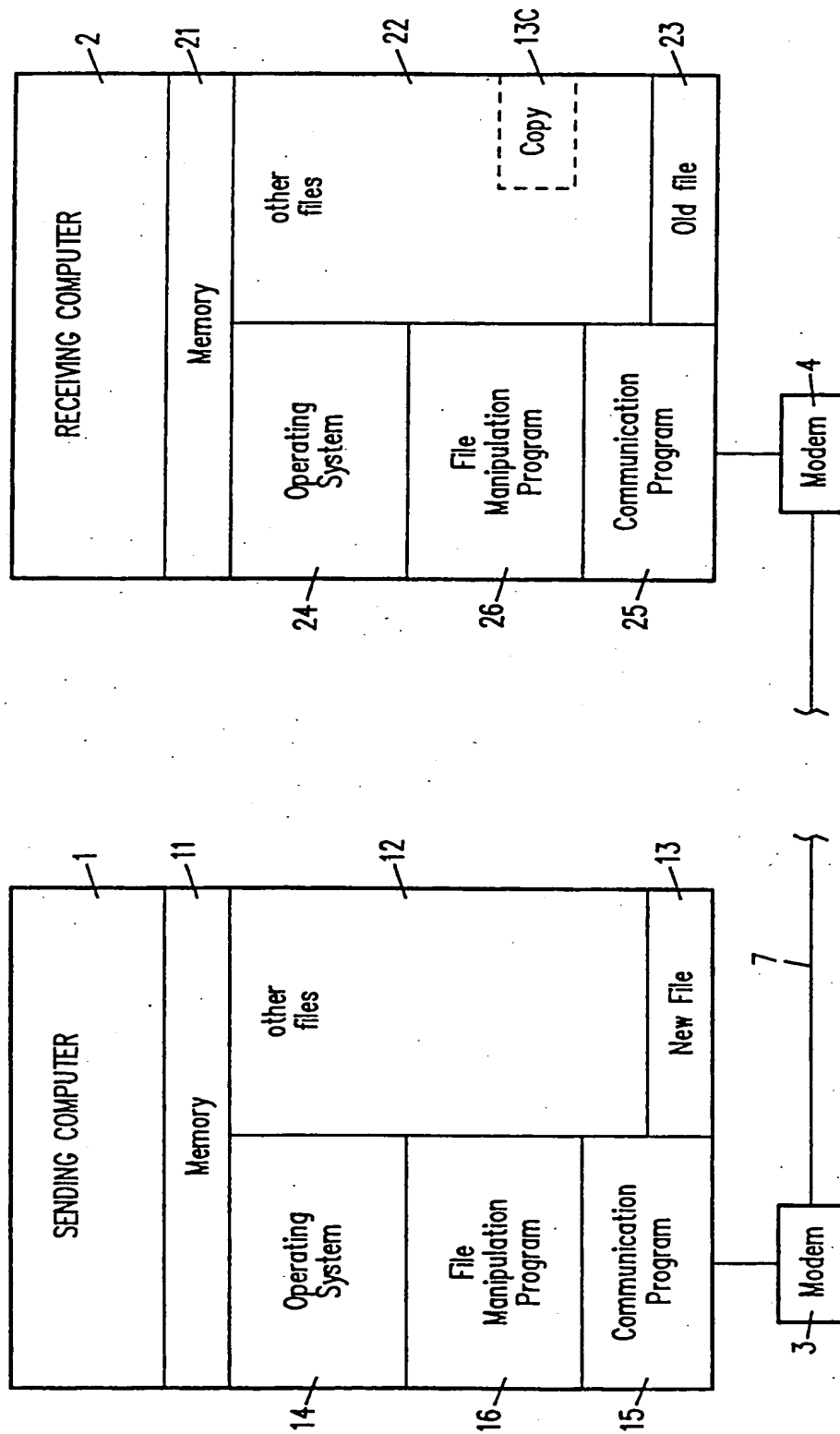


FIG. 1

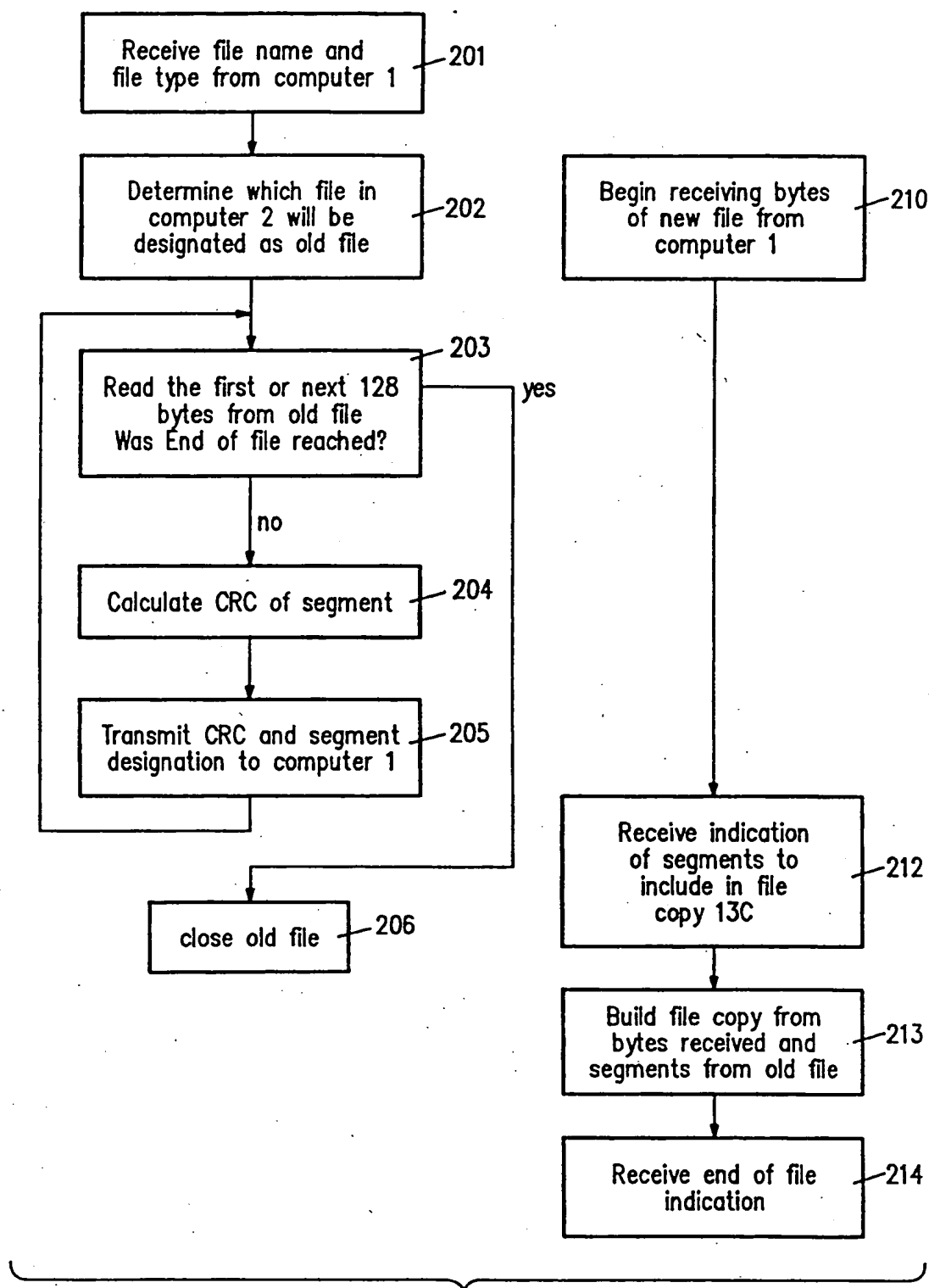


FIG. 2

| Segment in Old File | CRC of segment |
|---------------------|----------------|
| 0 | 435 |
| 1 | 529 |
| 2 | 314 |
| 3 | 435 |
| 4 | 529 |
| 5 | 314 |
| 6 | 435 |
| 7 | 529 |
| 8 | 314 |
| Other Segments | |

FIG. 3A

| Ref # | Information transmitted | Comment |
|-------|--|---|
| L1 | file name and file type of new file 13 | Allows computer 2 to pick which file to use as old file |
| L2 | Bytes 1-57 | No CRC match found |
| L3 | ID of segment 3 | CRC equals Segment 3 |
| L4 | Bytes 185-201 | No CRC match found |
| L5 | ID of segment 1 | CRC equals Segment 1 |
| L6 | Bytes 354-355 | No CRC match found |
| L7 | ID of segment 3 | Example of Repeat |
| L8 | Etc. | |

FIG. 3B

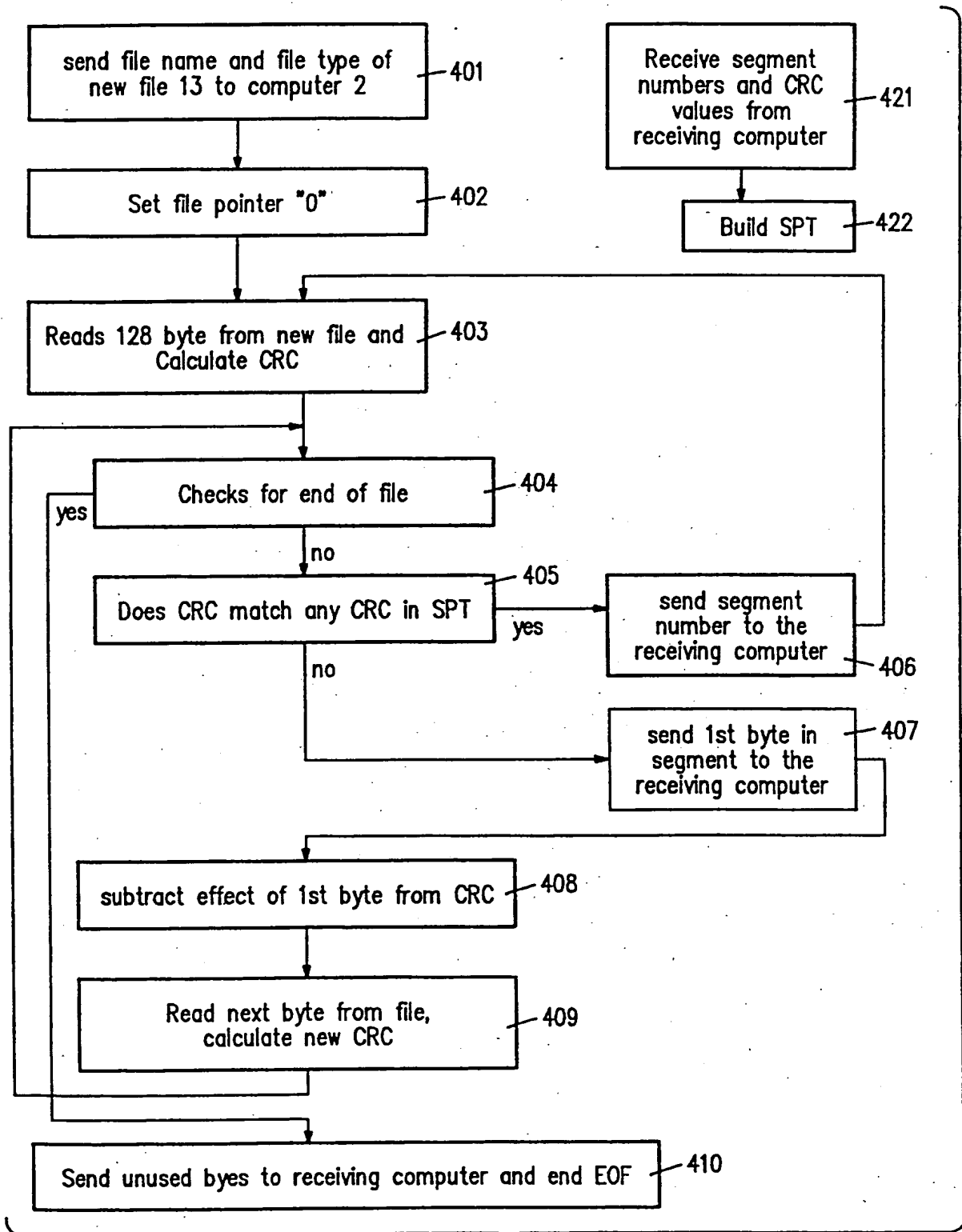


FIG. 4

4/4

SUBSTITUTE SHEET (RULE 26)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/14969

| A. CLASSIFICATION OF SUBJECT MATTER IPC(6) :GO6F 13/00 US CL :395/200 According to International Patent Classification (IPC) or to both national classification and IPC | | | | |
|--|--|---|--|--|
| B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) U.S. : 395/200, 600 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) APS search terms: cyclic redundancy check, hashing, file transfer, data compression | | | | |
| C. DOCUMENTS CONSIDERED TO BE RELEVANT | | | | |
| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. | | |
| Y | US, A, 4,641,274 (SWANK) 03 February 1987, column 3, lines 8-56, column 4, lines 19-47. | 1-10 | | |
| Y | US, A, 4,701,745 (WATERWORTH) 20 October 1987, ABSTRACT, column 1, lines 13-18. | 1-10 | | |
| <input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex. | | | | |
| <table border="0"> <tr> <td> * Special categories of cited documents: "A" document defining the general state of the art which is not considered to be part of particular relevance "E" earlier document published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed </td> <td> "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family </td> </tr> </table> | | | * Special categories of cited documents: "A" document defining the general state of the art which is not considered to be part of particular relevance "E" earlier document published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed | "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family |
| * Special categories of cited documents: "A" document defining the general state of the art which is not considered to be part of particular relevance "E" earlier document published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed | "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family | | | |
| Date of the actual completion of the international search 28 MARCH 1995 | | Date of mailing of the international search report 28 APR 1995 | | |
| Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-9564 | | Authorized officer <i>B. J. Ellcessor</i> LARRY J. ELLCESSOR Telephone No. (703) 305-3835 | | |